

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-098599

(43)Date of publication of application : 11.04.1995

(51)Int.Cl.

G10L 3/00

G10L 3/02

(21)Application number : 05-242764

(71)Applicant : NIPPON TELEGR &amp; TELEPH CORP &lt;NTT&gt;

(22)Date of filing : 29.09.1993

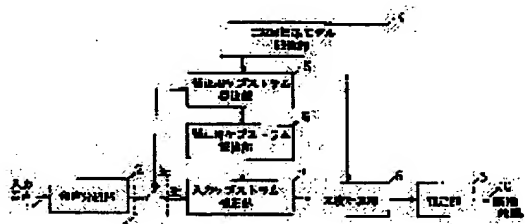
(72)Inventor : NAGASHIMA HIROMI  
TAKAHASHI JUNICHI  
HIDA MIZUHIRO

## (54) SOUND RECOGNIZING DEVICE

## (57)Abstract:

PURPOSE: To improve a recognizing factor of a voice signal distorted by a frequency characteristic of a transmission line such as voice in a telephone.

CONSTITUTION: In a transmission characteristic estimation mode, an input voice already know is converted to a time series of a feature vector consisting of a cepstrum coefficient,  $\Delta$  cepstrum coefficient and logarithmic power in an analyzing section 2, a group of a standard model of HMM indicating a voice of the same category as this input voice is taken out from a model storage section 4, a time series of these input voice and a state of the standard model are corresponded in a correction cepstrum calculating section 5, an average value of difference between a cepstrum coefficient of the time series of the input voice and a cepstrum coefficient of the state of the model is obtained, and stored in a storage section 6 as cepstrum for correction. Next, an unknown sound is inputted, converted to a time series of a feature vector in the same way in the analyzing section 2, a cepstrum coefficient is subtracted with cepstrum for correction of the storage section 6 and the result is corrected, the corrected time series of the feature vector is made a group of the standard model having high likelihood by a normal means in a likelihood calculating section 8 using the standard model 4, and it is made a recognized result.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the  
examiner's decision of rejection or application  
converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of  
rejection][Date of requesting appeal against examiner's decision  
of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-98599

(43)公開日 平成7年(1995)4月11日

(51)Int.Cl. <sup>4</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00	5 3 5	9379-5H		
3/02	3 0 1 A	9379-5H		

審査請求 未請求 請求項の数 2 O L (全 5 頁)

(21)出願番号 特願平5-242764

(22)出願日 平成5年(1993)9月29日

(71)出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72)発明者 長島 広海

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72)発明者 高橋 淳一

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72)発明者 飛田 瑞広

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(74)代理人 弁理士 草野 卓

(54)【発明の名称】 音声認識装置

(57)【要約】

【目的】 電話音声のように伝送路の周波数特性による変形を受けた音声信号の認識率を向上させる。

【構成】 伝送特性推定モードで、既知の入力音声进行分析部2でケプストラム係数、 $\Delta$ ケプストラム係数、対数パワーからなる特徴ベクトルの時系列に変換し、この入力音声と同一カテゴリの音声を表わすHMMの標準モデルの系列をモデル記憶部4から取出し、これら入力音声の時系列と、標準モデルの状態とを、補正ケプストラム算出部5で対応づけ、入力音声の時系列のケプストラム係数と、モデルの状態のケプストラム係数との差の平均を求め、補正用ケプストラムとして記憶部6に記憶する。次に未知音声を入力させ、分析部2で同様に特徴ベクトルの時系列に変換し、そのケプストラム係数から記憶部6の補正用ケプストラムを減算して補正し、この補正した特徴ベクトルの時系列を、記憶部4の標準モデルを用いて尤度計算部8で通常の手法で尤度の高い標準モデルの系列を認識結果とする。

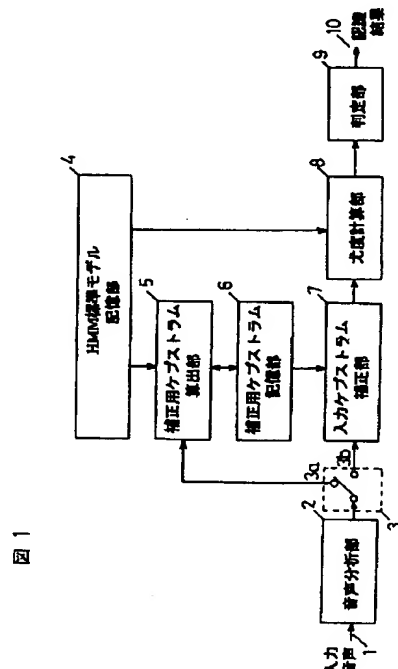


図 1

## 【特許請求の範囲】

【請求項1】 入力音声を音声分析手段によりケプストラム係数を含む特徴ベクトルの時系列に変換し、その時系列と標準モデル記憶部の標準モデルとの尤度を尤度計算手段で計算して、その最大の尤度が得られる標準モデルの音声認識結果として出力する音声認識装置において、

上記音声分析手段により分析された既知入力音声のケプストラム係数と上記標準モデル記憶部内の対応標準モデルのケプストラム係数との差を計算して補正用ケプストラムを作る補正用ケプストラム算出手段と、

上記補正用ケプストラムを記憶する補正用ケプストラム記憶部と、

上記音声分析手段により分析された未知入力音声の特徴ベクトルに対し、そのケプストラム係数から上記補正用ケプストラムを減算してその補正された特徴ベクトルを上記尤度計算手段へ供給する入力ケプストラム補正手段と、

を設けたことを特徴とする音声認識装置。

【請求項2】 入力音声を音声分析手段によりケプストラム係数を含む特徴ベクトルの時系列に変換し、その時系列と標準モデル記憶部の標準モデルとの尤度を計算してその最大の尤度が得られる標準モデルの音声認識結果として出力する音声認識装置において、

上記音声分析手段により分析された既知入力音声のケプストラム係数と上記標準モデル記憶部内の対応標準モデルのケプストラム係数との差を計算して補正用ケプストラムを作る補正用ケプストラム算出手段と、

上記補正用ケプストラムを上記標準モデル記憶部の各標準モデルのケプストラム係数に加算して標準モデルを補正するモデル補正手段と、

を設けたことを特徴とする音声認識装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 この発明は電話音声に代表されるような伝送特性によって周波数特性の変形を受けた音声を、あらかじめ推定した逆の周波数特性で補正し、認識率の向上を計った音声認識装置に関する。

## 【0002】

【従来の技術】 従来において、電話の送話器や回線などの周波数特性が多様に変わる伝送特性を経由した音声を認識させようとした場合には、種々の電話機と回線経路の組み合わせで音声を収録し、標準パターンあるいは隠れマルコフモデル（以下HMMと略する）の標準モデルを作成するための学習サンプルとする方法が通常おこなわれている。この方法によると学習サンプルの音声収録に多大な手間を必要とする。また、別の特性をもつ新しい電話機の出現への対応も困難である。さらに、このようにして収録された学習サンプルを用いて作成された標準パターンや標準モデルは、種々の伝送周波数特性をカ

2

バーするためになまったスペクトルの概形をもつことになり、ある程度の認識率の低下は避けられない。

【0003】 一方、伝送特性の影響をほぼ完全に除去する方法として入力音声ごとに自己の平均スペクトルで逆フィルタリングをおこなう方法が提案されている（日本音響学会講演論文集3-1-20 市川、中島、“話者認識における伝送系の影響”昭和52年10月）。この方法によると、短い音声や含まれる音節に片寄りのある音声の場合、音声認識に必要な音韻のスペクトルも平坦化されてしまうため、話者認識には有効であっても音声認識には適さない。

## 【0004】

【発明が解決しようとする課題】 この発明の目的は、単一の平坦な伝送特性の学習サンプルから作成した標準モデルでも電話音声のように周波数成分が制限された音声に対し安定した高い認識率が得られる音声認識装置を提供することにある。

## 【0005】

【課題を解決するための手段】 請求項1の発明によれば発声内容が既知の入力音声の特徴パラメータの時系列とそれに対応した標準モデルとのマッチングをおこなって、つまり差をとって、伝送特性の差をあらわす補正用ケプストラムを計算する補正用ケプストラム算出手段と、その結果の補正用ケプストラムを蓄えておく補正用ケプストラム記憶部と、補正用ケプストラムを認識されるべき入力音声のケプストラム係数から減算する入力ケプストラム補正手段とを設け、その補正した入力音声特徴ベクトルに対して認識処理を行う。

【0006】 請求項2の発明によれば請求項1の発明と同様に補正用ケプストラムを算出するが、請求項2の発明では入力音声ではなく標準モデルのそれぞれに対して、補正用ケプストラムを加算して補正するモデル補正手段を設け、その補正された標準モデルを用いて認識処理を行う。

## 【0007】

【作用】 ケプストラムが対数スペクトルのフーリエ変換に相当していることから、ケプストラムの加算はスペクトル上でフィルタをかけることに、ケプストラムの減算は逆フィルタをかけることに相当する。したがって、補正用ケプストラムが伝送特性を正しく近似していれば、入力ケプストラム補正手段の減算あるいはモデル補正手段の加算によって伝送特性の補正ができる。

## 【0008】

【実施例】 図1に請求項1の発明の第一の実施例を示す。最初に、入力音声から伝送周波数特性を推定するときの動作について説明する。入力端1から入力された既知の音声は、音声分析部2でLPCケプストラム係数、 $\Delta$ LPCケプストラム係数、 $\Delta$ 対数パワーからなる特徴ベクトルの時系列に変換され、この特徴ベクトルの時系列は切り換えスイッチ3の固定接点3a側を通じて補正

用ケプストラム算出部 5 に送られる。標準モデル記憶部 4 からは入力音声と同一カテゴリの音声をあらわす HMM の標準モデルの系列が補正用ケプストラム算出部 5 に送られる。補正用ケプストラム算出部 5 では、これらの 2 つの入力をもとにして以下のように補正用ケプストラムを計算する。まず、公知の方法であるビタビ (V i t e r b i) アルゴリズム (中川, “確率モデルによる音声認識”, p p. 44-46, 電子情報通信学会) によって入力音声の時系列の各時刻と標準モデルの状態との対応付けをもとめる。対応付けの一例を図 2 に示す。この例では、入力音声は / t o r i a t u k a u / という音声で、図 3 に示すような状態数が 3 の標準モデルを用いている。入力音声の時刻 1 では標準モデル t で状態 1 と対応し、時刻 2 でもモデル t の状態 1 と対応し、時刻 3 でモデル t の状態 2 が対応し、時刻 4 でもモデル t の状態 2 と対応し、時刻 5 でモデル t の状態 3 と対応し、時刻 6 でモデル 0 の状態 1 と対応する。以下同様に対応する。

【0009】次に、入力音声の時系列の各時刻のケプストラム係数と対応付けられた標準モデルの状態のケプストラム係数との差のケプストラム係数を計算する。この差のケプストラム係数を全時刻に渡って加算平均し、平均値をもとめて補正用ケプストラムとして補正用ケプストラム記憶部 6 に蓄えられる。標準モデルは、高品質音声信号データ、つまり周波数帯域幅がおよそ 100 Hz から 7~8 kHz であり、周波数特性が平坦でひずみがなく、SN 比が 30~40 dB 以上の条件で収録された音声信号データから作られている。従って前記のようにしてもとめられた補正用ケプストラムは間接的に伝送周波数特性を表現している。

【0010】次に、認識をおこなうときの動作について説明する。切り換えスイッチ 3 は固定接点 3 b 側に接続する。入力端 1 から入力された音声は、推定のときと同様に音声分析部 2 で L P C ケプストラム係数、 $\Delta$  L P C ケプストラム係数、 $\Delta$  対数パワーからなる特徴ベクトルの時系列に変換され、固定接点 3 b を通じて入力ケプストラム補正部 7 に送られる。入力ケプストラム補正部 7 では、補正用ケプストラム記憶部 6 に蓄えられた補正用ケプストラムを読み出し、入力音声のケプストラム係数との減算をおこなう。補正用ケプストラムの減算は伝送周波数特性の逆フィルタをかけることに相当する。特徴ベクトルの他の成分である  $\Delta$  L P C ケプストラム係数と  $\Delta$  対数パワーは伝送周波数特性の影響を本質的に受けていないと考えられるので、特に補正をおこなう必要はない。このようにして補正された入力音声の特徴ベクトルは、次に尤度計算部 8 に送られる。尤度計算部 8 では、認識対象の音声として可能性のあるすべての HMM の標準モデルの組み合わせについて、前記のビタビ (V i t

e r b i) アルゴリズムを用いて尤度を計算する。判定部 9 では、すべての組み合わせの尤度の中で最大値を選択し、それに対応する標準モデルの系列を認識結果として出力端 10 に出力する。

【0011】図 4 に請求項 2 の発明の第二の実施例を示し、図 1 と対応する部分に同一符号を付してある。これは図 1 の第一の実施例に入力ケプストラム記憶部 11 が追加されている。第一の実施例では、認識のモードと推定のモードとを明確に分けて、推定のモードであらかじめ決められた音声だけから補正用ケプストラムを計算していた。第二の実施例では、認識のモードで入力された音声は認識に使われるだけでなく、特徴ベクトルの時系列の形で入力ケプストラム記憶部 11 にも蓄えられ、認識結果が確定した時点で伝送周波数特性の推定のために、記憶部 11 から読み出して補正用ケプストラム算出部 5 に送られ、この入力音声の特徴ベクトル時系列と記憶部 4 の同一のカテゴリの HMM の標準モデルとの対応付けを同様に行って両者のケプストラム係数の差を求め、更に補正用ケプストラムを求め、認識のために入力された音声によっても補正用ケプストラムを少しずつ更新し、より正確な補正ができるようになる。

【0012】請求項 2 の発明の実施例を図 5 に示し、図 1 と対応する部分に同一符号を付けてある。この実施例でも入力既知音声からその伝送周波数特性を推定する際に、図 1 に示した場合と同様に、補正用ケプストラム算出部 5 で補正用ケプストラムを算出するが、この実施例では、その補正用ケプストラムをモデル補正部 13 で標準モデル記憶部 4 内の各標準モデルのケプストラム係数に加算して標準モデルを補正し、その補正した標準モデルを標準モデル記憶部 4 に記憶する。認識動作時には入力音声の特徴ベクトルを直接尤度計算部 8 へ供給して、前記補正された標準モデルを用いて認識処理を行う。

【0013】

【発明の効果】以上説明したようにこの発明によれば、内容のわかっている音声から伝送特性を推定し、入力音声に対してその伝送特性の補正をおこなうので、単一の平坦な伝送特性の学習サンプルから作成した標準モデルだけを用意しておくだけでも安定した高い認識率が得られる。

【図面の簡単な説明】

【図 1】この発明の第一の実施例を示すブロック図。

【図 2】入力音声の時系列の各時刻と標準モデルの状態との対応付けの一例を示す図。

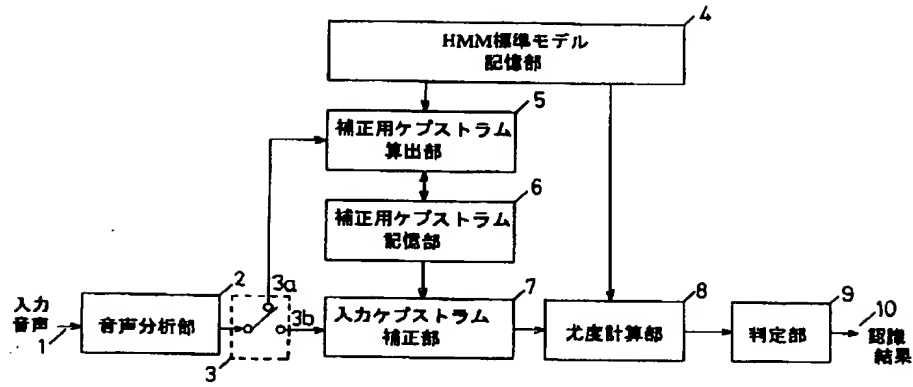
【図 3】HMM の標準モデルの状態遷移の一例を示す図。

【図 4】この発明の第二の実施例を示すブロック図。

【図 5】請求項 2 の発明の実施例を示すブロック図。

【図1】

図1



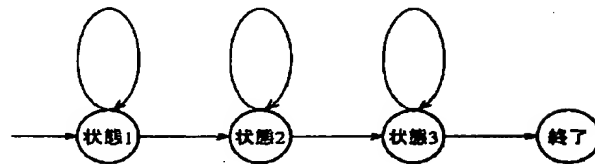
【図2】

図2

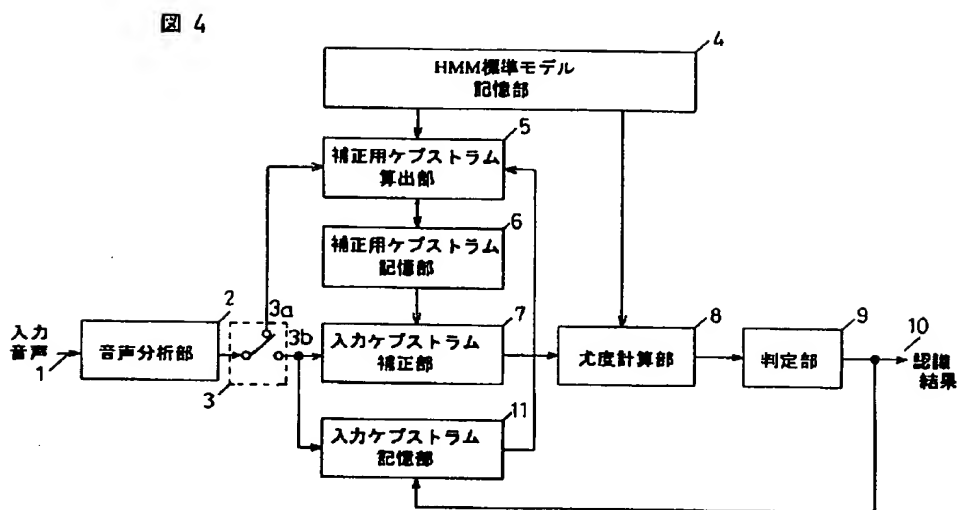
入力音声 の時刻	HMM 標準モデル	状態番号
1	t	1
2	t	1
3	t	2
4	t	2
5	t	3
6	o	1
7	o	1
8	o	1
9	o	2
10	o	2
11	o	2
12	o	3
13	o	3
14	r	1
15	r	1
.	.	.
.	.	.
.	.	.
.	.	.
154	a	3
155	a	3
156	u	1
157	u	2
158	u	2
159	u	2
160	u	3
161	u	3

【図3】

図3



【図 4】



【図 5】

